

1-1-2016

Metataxonomic and Metagenomic Approaches vs. Culture-Based Techniques for Clinical Pathology.

Sarah K. Hilton

George Washington University

Eduardo Castro-Nallar

George Washington University

Marcos Pérez-Losada

George Washington University

Ian Toma

George Washington University

Timothy A. McCaffrey

George Washington University

See next page for additional authors

Follow this and additional works at: https://hsrc.himmelfarb.gwu.edu/smhs_intsysbio_facpubs



Part of the [Medical Pathology Commons](#), [Pathology Commons](#), and the [Systems Biology Commons](#)

APA Citation

Hilton, S. K., Castro-Nallar, E., Pérez-Losada, M., Toma, I., McCaffrey, T. A., Hoffman, E. P., Siegel, M. O., Simon, G. L., Johnson, W. E., & Crandall, K. (2016). Metataxonomic and Metagenomic Approaches vs. Culture-Based Techniques for Clinical Pathology.. *Frontiers in Microbiology*, 7 (). <http://dx.doi.org/10.3389/fmicb.2016.00484>

This Journal Article is brought to you for free and open access by the Genomics and Precision Medicine at Health Sciences Research Commons. It has been accepted for inclusion in Genomics and Precision Medicine Faculty Publications by an authorized administrator of Health Sciences Research Commons. For more information, please contact hsrc@gwu.edu.

Authors

Sarah K. Hilton, Eduardo Castro-Nallar, Marcos Pérez-Losada, Ian Toma, Timothy A. McCaffrey, Eric P. Hoffman, Marc O. Siegel, Gary L. Simon, W. Evan Johnson, and Keith Crandall



Metataxonomic and Metagenomic Approaches vs. Culture-Based Techniques for Clinical Pathology

Sarah K. Hilton^{1*}, Eduardo Castro-Nallar^{1,2*}, Marcos Pérez-Losada^{1,3,4}, Ian Toma⁵, Timothy A. McCaffrey⁶, Eric P. Hoffman⁴, Marc O. Siegel⁷, Gary L. Simon⁷, W. Evan Johnson⁸ and Keith A. Crandall^{1*}

¹ Computational Biology Institute, The George Washington University, Ashburn, VA, USA, ² Facultad de Ciencias Biológicas, Center for Bioinformatics and Integrative Biology, Universidad Andres Bello, Santiago, Chile, ³ Centro de Investigación em Biodiversidade e Recursos Genéticos (CIBIO-InBIO), Vairão, Portugal, ⁴ Children's National Medical Research Center, Washington DC, USA, ⁵ Division of Genomic Medicine, Department of Medicine, The George Washington University School of Medicine and Health Sciences, Washington DC, USA, ⁶ Division of Genomic Medicine, Department of Medicine, Department of Microbiology, Immunology, and Tropical Medicine, The George Washington University School of Medicine and Health Sciences, Washington DC, USA, ⁷ Division of Infectious Diseases, Department of Medicine, School of Medicine and Health Sciences, The George Washington University, Washington DC, USA, ⁸ Computational Biomedicine, Boston University School of Medicine, Boston, MA, USA

OPEN ACCESS

Edited by:

Leonard Peruski,
US Centers for Disease Control and
Prevention, USA

Reviewed by:

Pieter De Maayer,
University of Pretoria, South Africa
Jozsef Soki,
University of Szeged, Hungary
Tianyu Zhang,
Chinese Academy of Sciences, China

*Correspondence:

Sarah K. Hilton
shilton@gwmail.gwu.edu;
Eduardo Castro-Nallar
eduardo.castro@unab.cl;
Keith A. Crandall
kcrandall@gwu.edu

Specialty section:

This article was submitted to
Infectious Diseases,
a section of the journal
Frontiers in Microbiology

Received: 04 January 2016

Accepted: 22 March 2016

Published: 07 April 2016

Citation:

Hilton SK, Castro-Nallar E,
Pérez-Losada M, Toma I,
McCaffrey TA, Hoffman EP, Siegel MO,
Simon GL, Johnson WE and
Crandall KA (2016) Metataxonomic
and Metagenomic Approaches vs.
Culture-Based Techniques for Clinical
Pathology. *Front. Microbiol.* 7:484.
doi: 10.3389/fmicb.2016.00484

Diagnoses that are both timely and accurate are critically important for patients with life-threatening or drug resistant infections. Technological improvements in High-Throughput Sequencing (HTS) have led to its use in pathogen detection and its application in clinical diagnoses of infectious diseases. The present study compares two HTS methods, 16S rRNA marker gene sequencing (metataxonomics) and whole metagenomic shotgun sequencing (metagenomics), in their respective abilities to match the same diagnosis as traditional culture methods (culture inference) for patients with ventilator associated pneumonia (VAP). The metagenomic analysis was able to produce the same diagnosis as culture methods at the species-level for five of the six samples, while the metataxonomic analysis was only able to produce results with the same species-level identification as culture for two of the six samples. These results indicate that metagenomic analyses have the accuracy needed for a clinical diagnostic tool, but full integration in diagnostic protocols is contingent on technological improvements to decrease turnaround time and lower costs.

Keywords: microbiome, metagenomics, metataxonomics, high throughput sequencing, drug resistance, pathogen detection

INTRODUCTION

Infectious diseases remain a significant health care burden in both the United States and worldwide. In 2011, infectious diseases accounted for ~4.0 million emergency department visits (Ambulatory and Hospital Care Statistics, 2016)¹, 3.9 million outpatient department visits (Ambulatory and Hospital Care Statistics, 2016)¹ and ~4% of patients in acute care hospitals develop a health care associated infection (Magill et al., 2014). The accurate identification of the pathogen(s) causing disease is crucial to the correct diagnosis and treatment for the infection. A comprehensive, accurate, and rapid diagnosis, including pathogen identification at the species level and antibiotic

¹ Available online at: <http://www.cdc.gov/nchs/ahcd.htm>

resistance pattern, enables physicians to use more targeted antimicrobial therapies for these patients (Didelot et al., 2012). Currently, there are many methods that can be used to characterize the microbial composition of a sample from an infected patient and identify the potentially causative agent(s) including culture, polymerase chain reaction (PCR), microarrays, and High-Throughput Sequencing (HTS) (Table 1).

Microbial culture has been considered the gold standard of diagnostic techniques for bacterial and fungal species and is widely used in clinical laboratories. Culture methods involve growing the pathogen on appropriate media, a method developed more than 100 years ago (Didelot et al., 2012). Further identification of the pathogen, especially to the species level, often requires biochemical tests. Additional tests, such as antibiotic resistance tests, may require additional cultures. Due to the limitations of the media utilized for growth, there will be an inherent bias to cultures. Culture methods can only confirm the presence of a microorganism that can grow on the selected media. Therefore, culture techniques may not be effective at identifying the presence of novel pathogens or known but unculturable pathogens (e.g., environmental and clinical isolates; Didelot et al., 2012). Nevertheless, culture remains the most widely used diagnostic method based on its extensive validation and cost-effectiveness (Köser et al., 2012).

Traditional PCR and, more recently, microarray assays analyze a pathogen's genetic profile, rather than the morphological, phenotypic, and biochemical features that standard culture methods utilize. PCR assays are quick, specific, and cheap when the list of possible targets is short. PCR's weakness as a diagnostic tool stems from its potential bias, because target sequences for primer design must be chosen before testing begins. Further, PCR has relatively low-throughput capabilities, because multiple PCR amplifications per sample negatively impact the cost- and time effectiveness of the method (Wang et al., 2006; McLoughlin, 2011).

High-density microarrays are able to detect thousands of pathogens simultaneously through the design of specific, degenerate, and tiled oligonucleotide probes. However, these probes need to be periodically updated as new genomes are sequenced, which is a time consuming and expensive process. Universal microarrays have been attempted, but exhibit bias as probe or target secondary structure can influence hybridization in an unpredictable pattern (Yang and Rothman, 2004; McLoughlin, 2011). PCR and microarrays are both able to leverage genetic differences in pathogens for greater resolution than culture methods, but the inherent bias found in both PCR and microarrays still presents an obstacle to achieve an efficient and agnostic diagnostic tool.

HTS is the most in-depth and unbiased method of obtaining genomic or metagenomic information (Metzker, 2009). Unlike PCR or microarrays, it does not require primer or probe design, it can be easily multiplexed, and the specificity and selectivity of the sequencing can be adjusted computationally after acquiring the data (Adams et al., 2009; Dunne et al., 2012). The two main methods of pathogen identification using HTS are marker sequencing, chiefly 16S ribosomal RNA (16S) but also internal transcribed spacer (ITS) region for fungal species,

and whole metagenomic shotgun sequencing. We adopt the nomenclatural suggestions of Marchesi and Ravel (Marchesi and Ravel, 2015) and refer to the high throughput marker (e.g., 16S or ITS) based approach to microbial diversity characterization as *metataxonomics* and the shotgun genomic sequencing approach as *metagenomics*. A common weakness for HTS methods is database bias, but as sequencing becomes cheaper and more widely used, databases are growing in size and diversity. For example, in 2014, the FDA GenomeTrakr project uploaded to NCBI an average of 848 *Salmonella* and *Listeria* genomes per month (Allard et al., 2016).

Metataxonomics using 16S sequencing is a widely used technique that relies on the conserved and variable regions of the bacterial 16S rRNA gene to make community-wide taxonomic classifications. As HTS technologies have improved, the read length and overall quality of the sequencing has also improved, allowing for greater species resolution (Klindworth et al., 2012). 16S sequencing is the most widely used technique for microbial diversity analysis and has been used to investigate various environments, from soil in Antarctica (Chong et al., 2012) to the human gut (Dethlefsen et al., 2008). Because the 16S HTS approach is a PCR-based approach, it suffers from the same issues described above for PCR. 16S sequencing only uses data from one multicopy gene, any two organisms with the same 16S rRNA gene sequence might be classified as the same strain under a 16S analysis, even if they were, in reality, different strains. For example, based on 16S sequences, *Escherichia coli* O157:H7 cannot be differentiated from *E. coli* K-12 (Weinstock, 2012) nor *Shigella flexneri* from *E. coli*. The 16S gene has been shown to have intra-organismal differences, such as multiple copies (Rajendhran and Gunasekaran, 2011) and intra-genomic heterogeneity (Rajendhran and Gunasekaran, 2011), which will negatively influence the method's accuracy.

Metagenomic sequencing avoids PCR bias and it is not restricted to only bacterial sequences. In addition, the coverage of the genome outside of the small 16S rRNA gene region means that specific, strain level discrimination is achievable. This has been shown in the metagenomic sequencing of cholera (Chin et al., 2011), tuberculosis (Garday et al., 2011), *E. coli* (Rasko et al., 2011), and methicillin-resistant *Staphylococcus aureus* (MRSA) (Köser et al., 2012). Currently, the high cost of metagenomic sequencing and the noisy signal due to host contamination are the greatest drawbacks of this approach. Metagenomic sequencing captures not only the pathogenic sequences, but also the human host's genetic material, which can overwhelm the signal from the pathogens (Kuczyński et al., 2011) and lead to an inaccurate classification of the pathogenic community. On the other hand, the human host genetic sequences can be an advantage in examining a genetic response to infection (Perez-Losada et al., 2015). However, metagenomic sequencing is also much more expensive than 16S sequencing, especially to achieve the coverage and depth needed for species identification (Quail et al., 2012).

Metataxonomics and metagenomics, with their culture independence and wealth of data, both have the potential to improve diagnostics. But before either one of these methods can

TABLE 1 | Taxonomic identification methods comparisons.

Technique	Speed (days)	Cost	Accuracy	Resistance mutations	Multispecies	High-throughput	Emergent pathogens	Reference update
Culture	2–14	\$	Genus or Species	No, additional assays	No	No	No	No
PCR	2	\$	Genus or Species	No, additional assays	No, additional assays	No	No	No
Microarray	2	\$	Genus or Species	Yes	Yes	Yes	No	Annually but costly
16S	1.5–2	\$\$	Genus or Species	No	Yes	Yes	Yes	Every run
Metagenomic	2–3	\$\$	Species or Strain	Yes	Yes	Yes	Yes	Every run

become fully integrated into diagnostic protocols, their relative benefits need to be compared and validated by culture methods (Dunne et al., 2012). In this study, 16S and genomic DNA shotgun sequencing are compared in their respective abilities to match the clinical, culture diagnosis of intubated patients being clinically assessed for possible ventilator associated pneumonia (VAP). Patients with VAP have a crude mortality rate of 20–70% and an attributable mortality rate of 10–40% (Heyland et al., 1999; Ashraf and Ostrosky-Zeichner, 2012; Luyt et al., 2013). A faster and more accurate pathogen diagnosis would hopefully lead to targeted antimicrobial therapy, thereby reducing the excessive use of broad-spectrum antibiotics, their antibiotic-associated side effects and healthcare costs (Aryee and Price, 2014) and possibly the mortality rate in these patients (Dupont et al., 2001).

MATERIALS AND METHODS

In order to evaluate the ability of HTS to match the culture inference, bronchial aspirate samples were taken from eight intubated patients from The George Washington University Hospital with suspected VAP. Three methods—16S sampling, metagenomic sampling, and traditional culturing—were employed to determine the infectious agent (Table 2). The collection of discarded aspirate samples for bacterial sequencing and de-identified clinical and microbiological data was approved by the GWU Institutional Review Board. The indication for endotracheal suctioning was solely based on the clinical evaluation of the patient's attending physician, as was the decision regarding any antibiotic therapy.

DNA Preparation for HTS

We followed the DNA extraction protocol of Toma et al. (2014). Illumina's Nextera DNA Sample Preparation Kit was used to generate and barcode the sequencing libraries for the genomic DNA (gDNA). The gDNA was sequenced on a HiSeq 2500 (100 bp; single-end reads; NCBI accession number SRP045601). Six of the eight samples were sequenced twice and the sequence results were combined into one fastq file. Six of the samples corresponded to the ones described in Toma et al. (2014) and full-length PCR-amplified 16S sequences using PacBio sequencing were obtained from NCBI under accession numbers SRP028704 and SRP031650 (Toma et al., 2014). We acknowledge the sample size is limited, yet these were the only samples available for follow up DNA work.

Culture

The deep endotracheal aspirates were submitted to the GWU Hospital microbiology laboratory for routine Gram-staining and microbial culture as described (Toma et al., 2014). In short, the most purulent or blood-tinged portions were used for a Gram-stain and bacterial culture on sheep blood, chocolate and MacConkey agars. The cultures on sheep blood and chocolate agars were incubated in 5% CO₂ at 35°C for at least 48 h, while the cultures on MacConkey agar were incubated in a non-CO₂ atmosphere at 35°C for at least 24 h. Significant growth was defined as moderate to heavy growth of an isolate in the second, third, or fourth quadrants of each plate. Organisms identification and susceptibility results were accomplished using the Vitek[®] 2 identification (ID) and antibiotic susceptibility testing (AST) cards (bioMérieux Marcy l'Etoile, France) following the standard operating procedures utilized by the GWU Hospital microbiology laboratory. The residual aspirate samples were frozen at –80°C until processing for DNA extraction. We used the culture results as the “gold standard” against which we compared the metataxonomic and metagenomic results.

Quality Control for HTS Data

The raw metagenomic reads were preprocessed using PrinSeq-Lite v. 0.20.3 (filtering reads and trimming 3' and 5' bases < 25 PHRED, removing exact duplicates, reads with undetermined bases, and low complexity reads using Dust filter = 30) (Schmieder and Edwards, 2011). Human reads were filtered using Bowtie2 (Langmead and Salzberg, 2012) by mapping the metagenomic reads against the latest human genome reference (hg19) (Church et al., 2011). The raw 16S reads were processed through the PacBio SmartPortal pipeline to filter out reads shorter than 100 bp, reads with no insert, low complexity or low quality reads and to trim adaptor sequences as described in Toma et al. (2014).

Taxonomic Profiling with 16S and Metagenomic Data

We used PathoLib from PathoScope 2.0 (Hong et al., 2014) to obtain all the sequence data under the bacterial, viral, and fungal taxonomy IDs from NCBI's nucleotide non-redundant (NCBI nt nr) database as of March 3, 2014 and downloaded “The All-Species Living Tree” Project (LTP) (Yarza et al., 2008; Munoz et al., 2011) 16S database, version 113. We mapped the metagenomic reads against the bacterial, viral, and fungal NCBI nt nr database and then against the hg19 and phix174 database

TABLE 2 | Culture, metagenomic, and metataxonomic (16S) inferences across samples.

Sample code	Metagenomic accession numbers	Culture inference	HTS inference			
			Metagenomic		Metataxonomic (16S)	
			Strain	ti number	Abundance	Strain
s002	SRX682947	Moderate <i>Acinetobacter baumannii</i>	<i>Acinetobacter baumannii</i> TDC-AB0715	980514	0.333328772	<i>Stenotrophomonas maltophilia</i>
			<i>Delftia acidovorans</i> SPH-1	398578	0.097767298	<i>Acinetobacter baumannii</i>
			<i>Escherichia coli</i> str. K-12 substr. MG1655	511145	0.071375325	<i>Acinetobacter denitrificans</i>
			<i>Acidovorax</i> sp. JS42	232721	0.057764467	
			Uncultured bacterium	77133	0.041203617	
s014	SRX682948	Moderate yeast	Human herpesvirus 1	10298	0.748475222	<i>Stenotrophomonas maltophilia</i>
			<i>Delftia acidovorans</i> SPH-1	398578	0.048150122	<i>Phocaeicola abscessus</i>
			<i>Escherichia coli</i> str. K-12 substr. MG1655	511145	0.036559297	<i>Acinetobacter denitrificans</i>
			<i>Candida parapsilosis</i>	5480	0.020277269	
			Uncultured bacterium	77133	0.017402208	
s017	SRX682949	Moderate <i>Klebsiella</i> , moderate yeast	<i>Candida glabrata</i>	5478	0.300566414	N/A
			<i>Streptococcus pasteurianus</i> ATCC 43144	981540	0.219646355	
			<i>Klebsiella pneumoniae</i>	573	0.087924829	
			<i>Delftia acidovorans</i> SPH-1	398578	0.074917391	
			<i>Acinetobacter xylosoxidans</i> A8	762376	0.041784424	
s043	SRX682950	Moderate <i>Pseudomonas aeruginosa</i>	<i>Pseudomonas aeruginosa</i> PAO1	208964	0.764117144	<i>Pseudomonas otitidis</i>
			<i>Pseudomonas aeruginosa</i>	287	0.059077537	<i>Gemella haemolysans</i>
			<i>Pseudomonas aeruginosa</i> DK2	1093787	0.054320167	<i>Staphylococcus aureus</i>
			<i>Acinetobacter baumannii</i> TDC-AB0715	980514	0.022756707	
			<i>Pseudomonas phage</i> DMS3	389469	0.021813898	
s049	SRX682951	Moderate <i>Pseudomonas aeruginosa</i>	<i>Pseudomonas aeruginosa</i> PAO1	208964	0.806335723	<i>Pseudomonas otitidis</i>
			<i>Pseudomonas aeruginosa</i>	287	0.132182547	<i>Azomonas agilis</i>
			<i>Delftia acidovorans</i> SPH-1	398578	0.015413769	<i>Acinetobacter baumannii</i>
			<i>Pseudomonas aeruginosa</i> LESB58	557722	0.006846147	
			<i>Pseudomonas aeruginosa</i> PA7	381754	0.005035477	
s070	SRX682952	Abundant oropharyngeal flora	<i>Neisseria meningitidis</i> MC58	122586	0.139880136	<i>Neisseria cinerea</i> ATCC 14685
			Uncultured bacterium	77133	0.122063293	<i>Streptococcus mitis</i>
			<i>Candida glabrata</i>	5478	0.071585191	<i>Corynebacterium pseudodiphtheriticum</i>
			<i>Streptococcus pasteurianus</i> ATCC 43144	981540	0.070621397	
			<i>Neisseria meningitidis</i> Z2491	122587	0.067880228	

(Continued)

TABLE 2 | Continued

Sample code	Metagenomic accession numbers	Culture inference	HTS inference				
			Metagenomic		Metataxonomic (16S)		
			Strain	ti number	Abundance	Strain	Abundance
s071	SRX682953	Abundant MRSA	<i>Staphylococcus aureus</i> subsp. <i>aureus</i> JH1	359787	0.877861749	<i>Staphylococcus aureus</i>	0.997915382
			<i>Staphylococcus epidermidis</i>	1282	0.04124046	<i>Enterococcus faecalis</i>	0.002038779
			<i>Enterococcus faecalis</i> V583	226185	0.014553905	<i>Raoultella ornithinolytica</i>	0.0000427
			<i>Staphylococcus epidermidis</i> ATCC 12228	176280	0.011848339		
			<i>Candida glabrata</i>	5478	0.010538713		
s074	SRX682954	Moderate GPCs, moderate <i>Streptococcus pneumoniae</i>	<i>Candida glabrata</i>	5478	0.453943438		N/A
			Uncultured fungus	175245	0.193800314		
			<i>Streptococcus pasteurianus</i> ATCC 43144	981540	0.079202375		
			<i>Pseudomonas aeruginosa</i> LESB58	557722	0.044692608		
			<i>Deiftia acidovorans</i> SPH-1	398578	0.036024421		

using the PathoScope 2.0 PathoMap module. The reads that mapped with a higher score to the hg19 or the phix174 database were removed. Similarly, we mapped the 16S sample reads against the LTP database and then against the hg19 and phix174 databases, removing the reads that mapped with a higher score to the hg19 or the phix174. We applied a Bayesian mixture read reassignment model (Francis et al., 2013), as implemented in the PathoID module of PathoScope 2.0, to both of the PathoMap mapping results. If a sample's results had multiple strains of the same species in the top hits, the PathoID module was run again with a new theta parameter of 10,000. In order to determine the presence or absence of the *mecA* gene (GenBank Accession number NX52593), we used BLAST+ (Camacho et al., 2009) to create a blast database from the reads of each sample's metagenomic sequences and queried the *mecA* gene sequence against each database using BLAST. All analyses were conducted on GWU's ColonialOne High-Performance computer cluster and computational time was recorded using the linux time(1) module.

Lab Validation of Fungal DNA

To confirm the presence of fungal DNA detected by the metagenomic analysis, we performed subsequent PCR using fungal ITS primers to verify inferred taxa from the original samples for which there was no 16S amplification. A PCR reaction for sequencing was performed using universal primers ITS1F and ITS4 targeting the nuclear ITS region (White et al., 1990) in samples s017 and s074. Amplification was performed in a 15 µL reaction volume consisting of 10.025 µL nuclease free water, 12.5 µL GoTaq green buffer (Promega, Madison, WI), 0.3 µL forward primer, 0.3 µL reverse primer, 0.3 µL dNTPs, 0.075 µL GoTaq (Promega, Madison, WI), and 1 µL of template, either gDNA from sample s017 or s074. PCR was performed using a Mastercycle Nexus Gradient PCR machine (Eppendorf). The thermal cycler was programmed for 2 min at 95°C for initial denaturation, followed by 30 cycles of 1 min at 95°C for denaturation, 1 min at 52°C for annealing, 3 min at 72°C for extension, and 10 min at 72°C for the final extension.

We cleaned the PCR products using ExoSAP (Affymetrix, Santa Clara, CA). Each ExoSAP reaction contained 1 µL PCR product and 2 µL of ExoSAP diluted 1:3, and the thermal cycler was programmed for 37°C for 15 min and 80°C for 15 min. Each product was sequenced in both directions and in duplicates. The cycle sequencing reaction contained 6.75 µL water, 0.5 µL primer (forward or reverse), 1.25 µL 5X buffer, 0.5 µL Big Dye (Life Technologies) mix, and 1 µL of the cleaned PCR product. The thermo cycler was programmed for 30 cycles of 95°C for 30 s, 50°C for 30 s, and 60°C for 4 min. ITS amplicons were sequenced on an ABI 3730XL platform.

We revised chromatograms, adjusted quality, and created the consensus sequences, one contig for sample s017 and two contigs for sample s074, in Geneious v. 7.1.5 (Kearse et al., 2012). These consensus sequences were aligned along with 32 other sequences from *Candida glabrata*, *C. albicans*, *C. dubliniensis*, *C. parapsilosis*, *C. rugosa*, *C. tropicalis*, and *Malassezia globosa* using MUSCLE alignment with 8 iterations (Supplementary Table 1). These species were chosen either

because they were identified by the metagenomic analysis (*C. glabrata*, *C. tropicalis*, *C. albicans*, and *M. globosa*) or because they are closely related to the top hit, *C. glabrata* (*C. dubliniensis* and *C. rugosa*). A maximum-likelihood tree was created using RAXML (Stamatakis, 2006) with 50 independent replicates and 10,000 bootstrap replicates. Analyses were performed using Geneious v. 7.1.5 (Kearse et al., 2012) and SumTrees (Guss et al., 2011).

RESULTS AND DISCUSSION

Our study resulted in three data sets. First, the 16S data (PacBio) which was collected previously (Toma et al., 2014) and raw data files with sequence from each patient were submitted in bas.h5 format to the NCBI Short Read Archive (SRA) under the accession numbers SRP028704 and SRP031650. Second, metagenomic data are new to this study and sequences were submitted in fastq format to NCBI SRA under the accession numbers SRX682947 to SRX682954. Third, the fungal ITS data are new to this study and those amplicon target sequences were deposited at NCBI under accession numbers KU936092–KU936095. Using these data, we found differences in the abilities of the metataxonomic approach vs. the metagenomic approach to identify pathogens characterized through culture techniques. Computational time also differed between approaches. Furthermore, the metagenomic approach identified fungal pathogens not captured in the metataxonomic approach or culture approach. These comparisons are detailed below.

The Effect of Sequencing Type on Computational Time

The mapping stage for the 16S analysis took an average of 7 min and 13 s and used an average of 50.65 CPU seconds in kernel. The mapping stage of the metagenomic analysis took an average of 23 h and 1 min and used an average of 80,427.72 CPU seconds in kernel, ~1500 times longer.

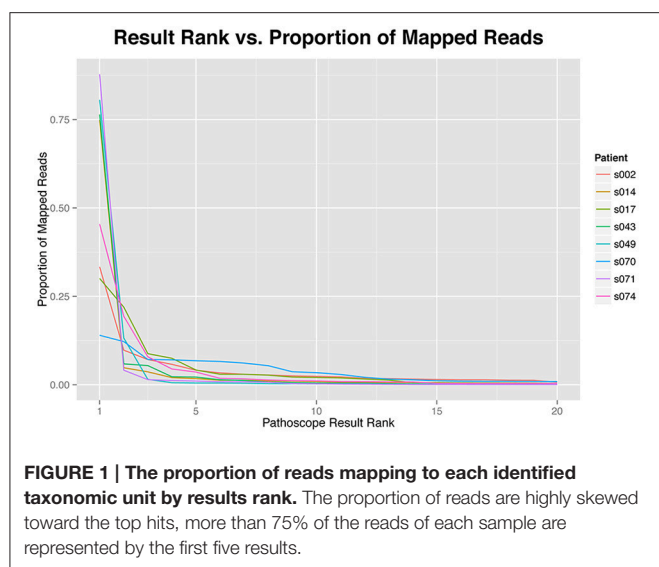
As predicted, the 16S analysis was much faster than the metagenomic analysis, due to the smaller number of sequences and the smaller database (**Supplementary Table 2**). The average 16S analysis took 5% of the time of the average metagenomic analysis and used only 0.1% of the CPU seconds. The turnaround time for each method is important because the integration of HTS into diagnostics hinges in part on whether or not a diagnosis can be produced as fast as or faster than current culture methods. A timely diagnosis allows clinicians to prescribe treatments quickly, which should improve the outcome of the treatment. Culture methods take on average 48 h to produce results (Didelot et al., 2012; Köser et al., 2012). Based on our results, the 16S analysis, at its most efficient, can match the 48 h turnaround timeline of culture, but the metagenomic analysis currently takes longer than the culture methods. The cost-effectiveness, efficiency, and coverage of metagenomic sequencing, and HTS in general, varies among different machines (Quail et al., 2012). However, if current trends continue, higher coverage with a shorter run time and lower cost should be achievable as the technology improves. While time and cost will not affect downstream analyses, an

increase in coverage will increase the already time-intensive mapping stage. Before a metagenomic analysis can be considered comparable to culture methods, improvements will have to be made in the mapping stage of the taxonomic profiling to decrease the amount of time required for metagenomic sequencing. It should be noted that the 48-h turnaround for culture methods is not true for all bacteria. Slow-growing bacteria or bacteria that require a series of cultures for diagnosis could take upwards of 2 weeks to complete the identification (Didelot et al., 2012). Likewise, fungal culture growth can take up to 1 week. Metagenomic sequencing does not depend on a culture step and all of the genetic information should, in theory, be collected through DNA extraction of the primary sample in a single step. Consequently, metataxonomic and metagenomic analyses are not affected by slow-growing or multi-culture pathogens and the turnaround times for these analyses have much smaller ranges than for the culture methods (Didelot et al., 2012).

Taxonomic Identification

The metataxonomic (16S) analysis identified an average of 8.5 operational taxonomic units (OTUs) per sample with a range of 3–18 OTUs. We designate OTUs as 16S (as mentioned in the introduction) often does not have resolving power for species and certainly strains due to the lack of sequence divergence at this one conserved locus (Caro-Quintero and Ochman, 2015). However, our approach does capitalize on the PacBio platform to achieve full-length amplicons of the 16S (~1500 bps) with greater than 100x coverage resulting in very high accuracy compared to many metataxonomic approaches that use short-read technology with lower coverage (Toma et al., 2014). The metagenomic analysis identified a larger number of OTUs with an average of ~374 OTUs per sample with a range of 185–797 OTUs. This high level of diversity is not unexpected; studies of the lung microbiome have found over 50 different bacterial genera in the lungs (Erb-Downward et al., 2011; Guss et al., 2011). For four of the samples, the metagenomic analysis identified several OTUs of the top hit species, while the 16S analysis did not contain any strain differentiation due to the lack of resolving power of this single locus approach. Multiple OTUs in the top hits could be indicative of multiple strains of the same species circulating in the sample or the presence of a novel species not found in the database. Samples s043 and s049 both had multiple OTUs in their top hits and the possibility of each scenario described above was explored using the PathoID theta parameter. The theta parameter controls the assumed number of ambiguous reads in a sample. Raising the theta parameter to 10,000 from the default assumes a more even distribution over several genomes in the database. A lack of change in the overall distribution between the results with the increased theta parameter vs. the default parameter indicates the top hit identified by PathoScope is most likely the correct strain in each sample (**Supplementary Table 3**).

While the increased diversity of hits in HTS analysis compared to PCR or microarrays may provide additional information, it has been shown that PathoScope and other Bayesian Mixture Models might create a tail of false positives (Lindner and Renard, 2015; Morfopoulou and Plagnol, 2015) (**Figure 1**). In the context of a clinical diagnosis, this tail of false positives should not influence



the final interpretation. We assume any pathogen causing an infection would be expected to constitute a large percentage of the pathogen population present in a sample and therefore would not be included in the tail of false positives. There are diseases for which this assumption will not hold, that is, the pathogen will not be the microbe in greatest abundance. This is especially true in non-sterile settings such as the gut, lung, nasal passage, where there is an abundance of microbes, some commensal, some opportunistic, and perhaps some pathogenic. Under these circumstances, it is essential to have appropriate controls for direct comparison of potentially pathogenic microbes that differ from “normal” microbiome components. With the metagenomic approach with RNAseq reads, one can also use the host data to examine host response to potential “infection” to help validate the identification of pathogens (Castro-Nallar et al., 2015).

The Effect of Sequencing on Clinical Diagnosis

The top hit from the metagenomic analysis was consistent with the culture inference at the species level in five of the six samples, while the culture inference and the 16S analysis top hit matched at the species level for only two of the six samples (Figure 2A). The two sequencing methods were only close in accuracy when the matches were considered at the genus level and the top five hits were taken into account (Figure 2D). The two samples that did not produce 16S amplification, s017 and s074, both had top hits of *C. glabrata*. This diagnosis is consistent with culture inference of s017 but not s074.

Sample s014 was the only sample that did not have the clinical diagnosis, i.e., yeast, matched as the top hit by either HTS method. However, the metagenomic analysis reported the fungus *C. parapsilosis* as the fourth hit with 2% of the reads. The 16S analysis never identified yeast, which is expected because the LTP database only contained bacterial sequences and 16S sequencing is restricted to bacteria. This is the only sample of the six samples analyzed by both HTS methods for which the

metagenomic analysis did not recover an organism that was identified by culture techniques.

The 16S and the metagenomic analyses were both able to match the culture inference for samples s070 and s071. The culture inference for sample s071 was MRSA and both sequencing analyses identified *S. aureus* as the top hit. However, the metagenomic analysis was able to provide additional information about a resistance profile that the 16S analysis could not. Sample s071 was the only sample for which metagenomic sequences mapped to the *mecA* gene at a significant abundance both in terms of breadth across the gene and depth in coverage (Figure 3). All 27 metagenomic sequences which mapped to the *mecA* gene during the BLAST search also map to the *S. aureus* JH1 genome during the metagenomic analysis, specifically to the SaurJH1_0029 locus which corresponds to the *mecA* gene. This gene has been shown to produce a penicillin binding protein, PBP 2', which has a low affinity for β -lactam antibiotics, including methicillin (Ubukata et al., 1989). Resistance profiles are incredibly important to clinicians and are critical for developing an appropriate antibiotic regimen. These results indicate it is possible to design sequencing based diagnostic tests which incorporate antibiotic resistance profiling along with taxonomic identification. The top hits of the 16S analysis and the metagenomic analysis of sample s070 corresponded to different species from the genus *Neisseria*. We consider this consistent with the culture inference of opportunistic flora. The top hit for the 16S analysis was *N. cinerea* and the top hit for the metagenomic analysis was *N. meningitidis*. *N. meningitidis* is found in the nasopharyngeal mucosa (Van Deuren et al., 2000) and can cause meningococcal pneumonia (Rose et al., 1981). *N. cinerea* is closely related to *N. gonorrhoeae*; while it is found in the oropharynx it has not been associated with lung infections (Knapp and Hook, 1988). The ambiguity of the culture inference does not allow for a clear distinction between the accuracy of 16S analysis and the accuracy of the metagenomic analysis. However, the metagenomic analysis does produce a top hit that is more likely to be the causative agent of pneumonia than 16S analysis top hit for sample s071.

The 16S and metagenomic analysis of samples s043 and s049 were consistent with the culture inference at the genus level, but the 16S analysis did not match the same species as the culture method. Both the metagenomic analysis and the culture approach detected the presence of *P. aeruginosa*, which is a known cause of pneumonia and often associated with nosocomial infections (Lister et al., 2009). The 16S analysis of both s043 and s049 identified *P. otitidis*, which is mainly associated with inner ear infections (Clark et al., 2006). The LTP database contains both *P. aeruginosa* and *P. otitidis*. However, the reassignment algorithm of PathoID (Francis et al., 2013) takes into account the uniqueness of the reads and for both samples there were more unique mappings to *P. otitidis* than *P. aeruginosa*. Therefore, PathoID reassigned all of the reads mapping to *P. otitidis* or *P. aeruginosa* to *P. otitidis*, resulting in the incorrect inference by the 16S analysis.

The final sample for which the 16S analysis failed to match the culture inference is sample s002. The clinical diagnosis, *Acinetobacter baumannii*, is the second hit for the 16S analysis,

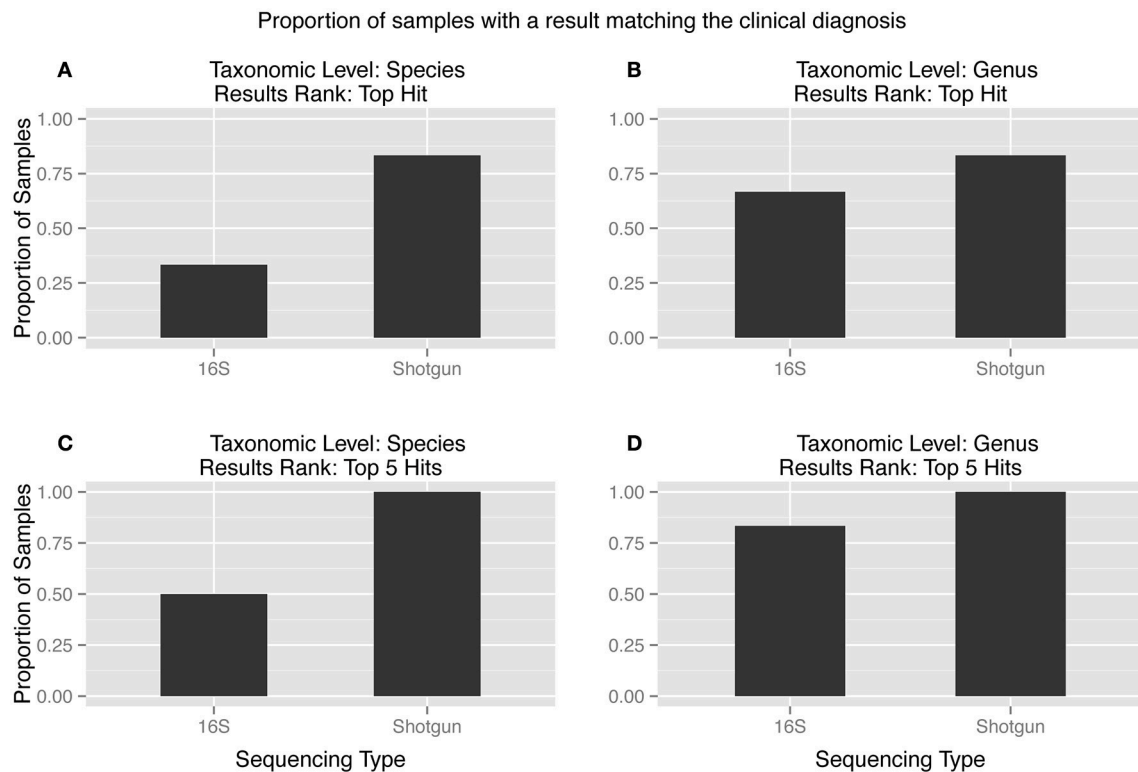


FIGURE 2 | (A–D): Proportion of samples matching the clinical diagnosis at different taxonomic levels and results rank. The metagenomic analysis is more accurate than the 16S analysis in every scenario. The ideal combination, a match between the clinical diagnosis and the tophit at the species level, showed the largest discrepancy between the two sequencing types.

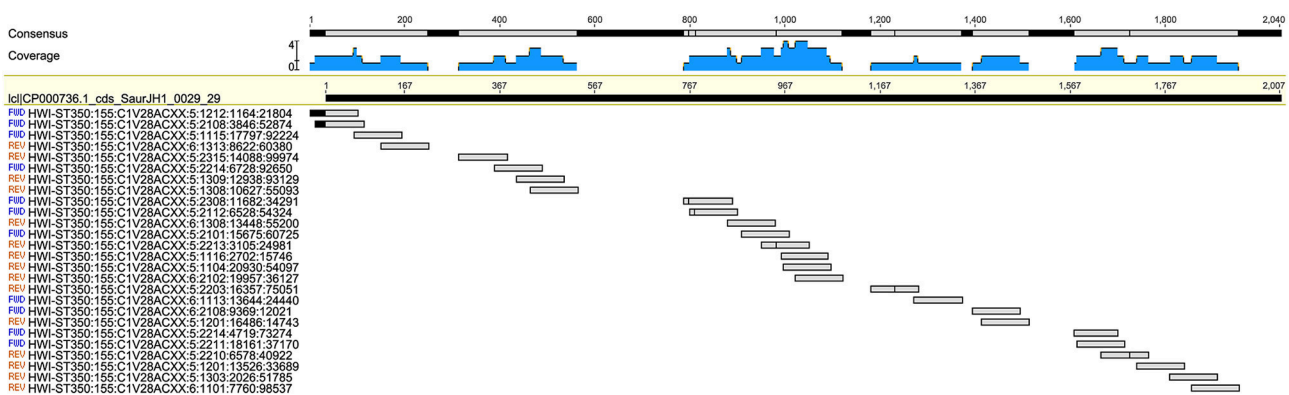


FIGURE 3 | Twenty-seven metagenomic reads from sample s070 mapping to the *mec-A* (penicillin-binding protein CDS) region of the *Staphylococcus aureus* genome.

but is only reported as having 16.7% of the reads while the top hit, *Stenotrophomonas maltophilia*, is reported as having 83% of the reads. However, *Stenotrophomonas* is a well-known pathogen in patients who develop VAP. Whether this was the true pathogen as opposed to *Acinetobacter* cannot be determined by the data at hand.

Samples s017 and s074 were not included in the comparisons of metataxonomic and metagenomic analysis results summarized

in **Figure 2** because the 16S PCR failed in both samples. The metagenomic sequencing was successful and the metagenomic analysis for both samples produced a top hit of *C. glabrata*, a fungus. The metagenomic sequencing inference was confirmed, in part, by the PCRs of the fungal ITS region of both samples. The PCR products cluster with the *C. glabrata* sequences in the phylogenetic tree with a bootstrap value of 0.83 (phylogeny not shown). The culture inference for sample s017 was yeast and

moderate *Klebsiella*; the metagenomic analysis has *C. glabrata* as the top hit and *K. pneumoniae* as the third hit. The culture inference of sample s074 was a moderate Gram-positive bacteria and moderate *S. pneumoniae*. The majority of the reads in s074 were *C. glabrata* and uncultured fungus, but the first bacterial identification, *S. pasteurianus*, matches the culture diagnosis at the genus level. The PCR test can only confirm the presence of *C. glabrata* in the sample and cannot confirm that *C. glabrata* is indeed the most prevalent organism. However, these results do show metagenomic sequencing is able to produce plausible results even when the 16S sequencing fails.

The Effect of Sequencing on Human Respiratory Pathogen Identification

HTS, with its many advantageous characteristics, including lack of specific primer/probe design and easy multiplexing which produces large amounts of sequence data, has been instrumental in investigating the human microbiome (Turnbaugh et al., 2007). Characterization of the microbiome provides the context of a “healthy” bacterial composition to studies that investigate diseases traditionally associated with an infectious component (Beck et al., 2012) and opens the door to understanding disease as an alteration in the overall microbial community, rather than as the invasion of one particular pathogen, exemplified by the metataxonomic study of *Clostridium difficile* infections (Chang et al., 2008).

The microbiome of the lung has been particularly difficult to characterize due to prior assumptions about the community composition of the lung, the diversity of pathogens causing diseases, and sampling concerns. Until very recently, the lung was considered to be a sterile body site and it was not included as a sampling site in the original Human Microbiome Project (Beck et al., 2012). Newer studies have shown there is indeed a microbial community within the lung, but it has been difficult to characterize this community. Some studies have failed to find a distinct lung microbiome (Charlson et al., 2011), while others have found the location of the sampling within the lung will produce different results (Erb-Downward et al., 2011), or have raised concerns about the level of contamination of the lung microbiome from other body sites, such as the oropharynx (Morris et al., 2013). While the resolution is not particularly clear, there are several genera that are proposed as possible core genera of the lung microbiome including *Pseudomonas*, *Streptococcus*, *Prevotella*, *Fusobacterium*, *Haemophilus*, *Veillonella*, and *Porphyromonas* (Morris et al., 2013). Patients with pneumonia have been found to have microbiomes dominated by pathogens such as *S. pneumoniae*, *H. influenzae*, and *K. pneumoniae* (Linder et al., 2014). The core genera of the microbiome proposed by Morris et al. can be found in the metagenomic analyses. However, not all of the genera are found in all of the samples, and they are found in small proportions unless they are a top hit, such as *Pseudomonas* in samples s043 and s049. The metataxonomic analysis produced an even sparser representation of these core genera than the metagenomic analysis. Of the pathogens that are said to dominate the microbiomes of patients with pneumonia, *K. pneumoniae* is

the only one found among the top hits of any sample. The results of our study have some similarities in genera with other lung microbiome studies, but these species are often found in small proportions. This may be a result of our samples coming from infected individuals, or may be because the lung microbiome is still in the process of being well-characterized.

While the results of our study were not completely congruent with other lung microbiome studies, the results did seem to correlate with respiratory pathogens. Of the top four hits of the metataxonomic and metagenomic analyses that would be considered part of the human flora or human pathogens rather than environmental pathogens, there were six organisms—*N. meningitidis* (Van Deuren et al., 2000), *Corynebacterium pseudodiphtheriticum* (Nhan et al., 2012), *K. pneumoniae* (Bratu et al., 2005), *A. baumannii* (Garnacho-Montero et al., 2003), *P. aeruginosa* (de Bentzmann et al., 1996; Lister et al., 2009), and *S. aureus* (Hooper and Smith, 2012) - that commonly cause respiratory infections. Of these respiratory pathogens, *A. baumannii*, *P. aeruginosa*, and *S. aureus* are commonly associated with VAP (Ashraf and Ostrosky-Zeichner, 2012). These respiratory pathogens were identified as the top hit more frequently by the metagenomic analysis than by the metataxonomic analysis. Out of the eight metagenomic analyses, five of the top hits are represented in this list of six respiratory pathogens. In the 16S analyses, only one of the six top hits fall into this category.

With only three exceptions, the top four hits in both the metataxonomic and metagenomic approaches identified microbes found in humans, either as part of the normal flora or as pathogens. These exceptions were found in both the metagenomic and the metataxonomic analyses, but the environmental microbes never accounted for more than 10% of the reads. The first exception, *Delftia acidovorans*, was identified by the metagenomic analysis in every sample except for s070 and s071. *D. acidovorans* is mainly found in soil, though it has been identified in the infection of a child with an endotracheal tube (Khan et al., 2012). The metagenomic analysis also identified *Acidovorax JS42* as the third hit for sample s002. *Acidovorax* is an environmental bacteria most often associated with plant infections, rather than human infections, though there has been one reported case of sepsis caused by *Acidovorax* (Willems et al., 1990; Shetty et al., 2005). Finally, the second hit in the 16S analysis of sample s049 was *Azomonas agilis*, an environmental bacterium that has neither been found in human flora nor been linked to an infection (Chebotar et al., 2001). These environmental microbes are not expected in the human lung. Their presence could be explained by contamination, the inhalation of soil particles by the patient or an improperly cleaned medical device, but a computational reason, such as an improperly mapped read, cannot be ruled out.

HTS beyond Primary Identification of Pathogens

Metagenomic analysis also presents opportunities to investigate a patient's infection that go beyond taxonomic identification of microbes. Metagenomic sequencing has been used on numerous

clinical examples of antibiotic resistance pathogens to identify new genes linked to antibiotic resistance and to characterize the distribution of these genes outside the clinical setting (Forsberg et al., 2012; Wright and Poinar, 2012). All of the resulting information is being pooled in many databases, such as the Comprehensive Antibiotic Resistance Database (McArthur et al., 2013). Dual gene expression uses RNA sequences to analyze the transcriptomes of the pathogen and the host during infection (Perez-Losada et al., 2015). This should shed light on new virulence factors in the pathogens and pathways activated in response to pathogens or pathogen-associated molecular patterns as the infection progresses (Chang et al., 2011; Westermann et al., 2012). These features of metagenomics provide added value and insights into infections and potential treatment options. Metagenomics could also be useful with infections of unknown etiology due to the immense amount of genetic information captured via an unbiased method (Wilson et al., 2014) or for creating strain or population specific primers (Rohde et al., 2011).

CONCLUSION

While our sample size is limited, our results indicate HTS has the potential to match standard culture techniques for the identification of bacterial pathogens. The two different HTS methods used for this study did have differences in their diagnostic capabilities and represent a tradeoff, namely between the speed of metataxonomic analyses and the accuracy and added information of metagenomic analyses. Metataxonomics is a quick, inexpensive choice when the pathogen of interest is a known bacterium, but if there are unknown pathogens, novel pathogens, a mixture of viruses, fungi and bacteria, our results show metagenomics performs better (e.g., our ability to identify fungal species with metagenomics). An effective diagnostic test must not only be fast and accurate, but must also be inexpensive, because hospitals and other healthcare facilities would be running numerous tests at anytime. Based on our results, neither HTS method can compete with the culture method in all three of these categories. However, HTS technology has seen rapid improvement over the last decade, and as these improvements continue (Shinshkin et al., 2015), HTS as a diagnostic tool will become faster and cheaper. Because diagnosis of infectious diseases is an important component of patient care, any improvements (e.g., speed, accuracy, and/or related features such as drug resistance profiles, human immune response profiles) to the diagnostic process would have an immediate impact for patients.

REFERENCES

- Adams, I. P., Glover, R. H., Monger, W. A., Mumford, R., Jackeviciene, E., Navalinskiene, M., et al. (2009). Next-generation sequencing and metagenomic analysis: a universal diagnostic tool in plant virology. *Mol. Plant Pathol.* 10, 537–545. doi: 10.1111/j.1364-3703.2009.00545.x
- Allard, M. W., Strain, E., Melka, D., Bunning, K., Musser, S. M., Brown, E. W., et al. (2016). The PRACTICAL value of Food Pathogen Traceability through

AUTHOR CONTRIBUTIONS

KC, EC, IT, TM conceived of the project; IT, TM, EH, MS, GS conducted clinical work associated with the project; IT, TM collected 16S data; WJ, SH, EC, MP, KC collected metagenomic data; SH, EC, MP, IT, WJ, KC conducted data analyses; SH, EC performed fungal PCR experiments; all authors contributed to results interpretation and all contributed to reviewing and writing of the manuscript.

FUNDING

SH: GW's Biology Department, Harlan Undergraduate Research Program, National Science Foundation's Research Experience for Undergraduates. EC: GW's Computational Biology Institute, "CONICYT + PAI/ CONCURSO NACIONAL APOYO AL RETORNO DE INVESTIGADORES/AS DESDE EL EXTRANJERO, CONVOCATORIA 2014 + FOLIO 82140008." MP: K12 Career Development Program 5 K12 HL119994 award. KC, GS, MS, IT, and TM: This project was supported by Award Number UL1TR000075 from the NIH National Center for Advancing Translational Sciences. Its contents are solely the responsibility of the authors and do not necessarily represent the official views of the National Center for Advancing Translational Sciences or the National Institutes of Health. Additional support was provided by the Abramson Family Trust.

ACKNOWLEDGMENTS

SH would like to thank Amy Zanne, Darcy Young, and David Stern for their help with the lab validation. SH and EC would like to thank GW's Colonial One High-Performance Computing facility for facilitating infrastructure for analyses.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fmicb.2016.00484>

Supplementary Table 1 | NCBI Accession numbers used in the phylogenetic tree and their associated species name.

Supplementary Table 2 | Database and run statistics for the 16S and the metagenomic analysis.

Supplementary Table 3 | Sample s043 and s049 metagenomic read proportions mapping to *Pseudomonas aeruginosa* strains under two different theta parameters.

- BUILDING a Whole-Genome Sequencing Network and database. *J. Clin. Microbiol.* doi: 10.1128/JCM.00081-16. [Epub ahead of print].
- Aryee, A., and Price, N. (2014). Antimicrobial Stewardship – can we afford to do without it? *Br. J. Clin. Pharmacol.* 79, 173–181. doi: 10.1111/bcp.12417
- Ashraf, M., and Ostrosky-Zeichner, L. (2012). Ventilator-associated pneumonia. *Hosp. Pract.* 40, 93–105. doi: 10.3810/hp.2012.02.950
- Beck, J. M., Young, V. B., and Huffnagle, G. B. (2012). The microbiome of the lung. *Trans. Res.* 160, 258–266. doi: 10.1016/j.trsl.2012.02.005

- Bratu, S., Landman, D., Haag, R., Recco, R., Eramo, A., Alam, M., et al. (2005). Rapid spread of carbapenem-resistant *Klebsiella pneumoniae* in New York city: a new threat to our antibiotic armamentarium. *Arch. Intern. Med.* 165, 1430–1435. doi: 10.1001/archinte.165.12.1430
- Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., et al. (2009). BLAST+: architecture and applications. *BMC Bioinformatics* 10:421. doi: 10.1186/1471-2105-10-421
- Caro-Quintero, A., and Ochman, H. (2015). Assessing the unseen bacterial diversity in microbial communities. *Genome Biol. Evol.* 7, 3416–3425. doi: 10.1093/gbe/evv234
- Castro-Nallar, E., Shen, Y., Freishtat, R. J., Perez-Losada, M., Manimaran, S., Liu, G., et al. (2015). Integrating microbial and host transcriptomics to characterize asthma-associated microbial communities. *BMC Med. Genomics* 8:50. doi: 10.1186/s12920-015-0121-1
- Chang, J. Y., Antonopoulos, D. A., Kalra, A., Tonelli, A., Khalife, W. T., Schmidt, T. M., et al. (2008). Decreased diversity of the fecal microbiome in recurrent *Clostridium difficile*—associated diarrhea. *J. Infect. Dis.* 197, 435–438. doi: 10.1086/525047
- Chang, S. T., Sova, P., Peng, X., Weiss, J., Law, G. L., Palermo, R. E., et al. (2011). Next-generation sequencing reveals HIV-1-mediated suppression of T cell activation and RNA processing and regulation of noncoding RNA expression in a CD4+ T cell line. *MBio* 2, e00134–e00111. doi: 10.1128/mBio.00134-11
- Charlson, E. S., Bittinger, K., Haas, A. R., Fitzgerald, A. S., Frank, I., Yadav, A., et al. (2011). Topographical continuity of bacterial populations in the healthy human respiratory tract. *Am. J. Respir. Crit. Care Med.* 184, 957–963. doi: 10.1164/rccm.201104-0655OC
- Chebota, V. K., Asis, C. A., and Akao, S. (2001). Production of growth-promoting substances and high colonization ability of rhizobacteria enhance the nitrogen fixation of soybean when coinoculated with *Bradyrhizobium japonicum*. *Biol. Fertil. Soils* 34, 427–432. doi: 10.1007/s00374-001-0426-4
- Chin, C.-S., Sorenson, J., Harris, J. B., Robins, W. P., Charles, R. C., Jean-Charles, R., et al. (2011). The origin of the Haitian cholera outbreak strain. *N. Engl. J. Med.* 364, 33–42. doi: 10.1056/NEJMoa1012928
- Chong, C., Pearce, D., Convey, P., Yew, W., and Tan, I. (2012). Patterns in the distribution of soil bacterial 16S rRNA gene sequences from different regions of Antarctica. *Geoderma* 181, 45–55. doi: 10.1016/j.geoderma.2012.02.017
- Church, D. M., Schneider, V. A., Graves, T., Auger, K., Cunningham, F., Bouk, N., et al. (2011). Modernizing reference genome assemblies. *PLoS Biol.* 9:e1001091. doi: 10.1371/journal.pbio.1001091
- Clark, L. L., Dajcs, J. J., Mclean, C. H., Bartell, J. G., and Stroman, D. W. (2006). *Pseudomonas otitidis* sp. nov., isolated from patients with otic infections. *Int. J. Syst. Evol. Microbiol.* 56, 709–714. doi: 10.1099/ijs.0.63753-0
- de Bentzmann, S., Plotkowski, C., and Puchelle, E. (1996). Receptors in the *Pseudomonas aeruginosa* adherence to injured and repairing airway epithelium. *Am. J. Respir. Crit. Care Med.* 154, S155–S162. doi: 10.1164/ajrcm/154.4_Pt_2.S155
- Dethlefsen, L., Huse, S., Sogin, M. L., and Relman, D. A. (2008). The pervasive effects of an antibiotic on the human gut microbiota, as revealed by deep 16S rRNA sequencing. *PLoS Biol.* 6:e280. doi: 10.1371/journal.pbio.0060280
- Didelot, X., Bowden, R., Wilson, D. J., Peto, T. E., and Crook, D. W. (2012). Transforming clinical microbiology with bacterial genome sequencing. *Nat. Rev. Genetics* 13, 601–612. doi: 10.1038/nrg3226
- Dunne, W. Jr., Westblade, L., and Ford, B. (2012). Next-generation and whole-genome sequencing in the diagnostic clinical microbiology laboratory. *Eur. J. Clin. Microbiol. Infect. Dis.* 31, 1719–1726. doi: 10.1007/s10096-012-1641-7
- Dupont, H., Mentec, H., Sollet, J., and Bleichner, G. (2001). Impact of appropriateness of initial antibiotic therapy on the outcome of ventilator-associated pneumonia. *Intensive Care Med.* 27, 355–362. doi: 10.1007/s001340000640
- Erb-Downward, J. R., Thompson, D. L., Han, M. K., Freeman, C. M., McCloskey, L., Schmidt, L. A., et al. (2011). Analysis of the lung microbiome in the “healthy” smoker and in COPD. *PLoS ONE* 6:e16384. doi: 10.1371/journal.pone.0016384
- Forsberg, K. J., Reyes, A., Wang, B., Selleck, E. M., Sommer, M. O., and Dantas, G. (2012). The shared antibiotic resistome of soil bacteria and human pathogens. *Science* 337, 1107–1111. doi: 10.1126/science.1220761
- Francis, O. E., Bendall, M., Manimaran, S., Hong, C., Clement, N. L., Castro-Nallar, E., et al. (2013). Pathoscope: species identification and strain attribution with unassembled sequencing data. *Genome Res.* 23, 1721–1729. doi: 10.1101/gr.150151.112
- Gardy, J. L., Johnston, J. C., Sui, S. J. H., Cook, V. J., Shah, L., Brodtkin, E., et al. (2011). Whole-genome sequencing and social-network analysis of a tuberculosis outbreak. *N. Engl. J. Med.* 364, 730–739. doi: 10.1056/NEJMoa1003176
- Garnacho-Montero, J., Ortiz-Leyba, C., Jimenez-Jimenez, F., Barrero-Almodovar, A., Garcia-Garmendia, J., Bernabeu-Wittell, M., et al. (2003). Treatment of multidrug-resistant *Acinetobacter baumannii* ventilator-associated pneumonia (VAP) with intravenous colistin: a comparison with imipenem-susceptible VAP. *Clin. Infect. Dis.* 36, 1111–1118. doi: 10.1086/374337
- Guss, A. M., Roeselers, G., Newton, I. L., Young, C. R., Klepac-Ceraj, V., Lory, S., et al. (2011). Phylogenetic and metabolic diversity of bacteria associated with cystic fibrosis. *ISME J.* 5, 20–29. doi: 10.1038/ismej.2010.88
- Heyland, D. K., Cook, D. J., Griffith, L., Keenan, S. P., and Brun-Buisson, C. (1999). The attributable morbidity and mortality of ventilator-associated pneumonia in the critically ill patient. *Am. J. Respir. Crit. Care Med.* 159, 1249–1256. doi: 10.1164/ajrcm.159.4.9807050
- Hong, C., Manimaran, S., Shen, Y., Perez-Rogers, J. F., Byrd, A. L., Castro-Nallar, E., et al. (2014). PathoScope 2.0: a complete computational framework for strain identification in environmental or clinical sequencing samples. *Microbiome* 2:33. doi: 10.1186/2049-2618-2-33
- Hooper, C. Y., and Smith, W. J. (2012). Telavancin for the treatment of nosocomial pneumonia caused by methicillin-resistant *Staphylococcus aureus* (MRSA). *Ther. Clin. Risk Manag.* 8, 131. doi: 10.2147/TCRM.S23247
- Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., et al. (2012). Geneious basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* 28, 1647–1649. doi: 10.1093/bioinformatics/bts199
- Khan, S., Sistla, S., Dhodapkar, R., and Parija, S. C. (2012). Fatal *Delftia acidovorans* infection in an immunocompetent patient with empyema. *Asian Pacific J. Trop. Biomed.* 2, 923–924. doi: 10.1016/S2221-1691(12)60254-8
- Klindworth, A., Pruesse, E., Schweer, T., Peplies, J., Quast, C., Horn, M., et al. (2012). Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. *Nucleic Acids Res.* 41:e1. doi: 10.1093/nar/gks808
- Knapp, J., and Hook, E. D. (1988). Prevalence and persistence of *Neisseria cinerea* and other *Neisseria* spp. in adults. *J. Clin. Microbiol.* 26, 896–900.
- Köser, C. U., Holden, M. T., Ellington, M. J., Cartwright, E. J., Brown, N. M., Ogilvy-Stuart, A. L., et al. (2012). Rapid whole-genome sequencing for investigation of a neonatal MRSA outbreak. *N. Engl. J. Med.* 366, 2267–2275. doi: 10.1056/NEJMoa1109910
- Kuczynski, J., Lauber, C. L., Walters, W. A., Parfrey, L. W., Clemente, J. C., Gevers, D., et al. (2011). Experimental and analytical tools for studying the human microbiome. *Nat. Rev. Genet.* 13, 47–58. doi: 10.1038/nrg3129
- Langmead, B., and Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9, 357–359. doi: 10.1038/nmeth.1923
- Linder, A., Cirstea, M., Hancock, R. E. W., Boyd, J. H., and Fjell, C. (2014). The microbiome of severe community acquired pneumonia. *Am. J. Respir. Crit. Care Med.* 189, A6715.
- Lindner, M. S., and Renard, B. Y. (2015). Metagenomic profiling of known and unknown microbes with MicrobeGPS. *PLoS ONE* 10:e0117711. doi: 10.1371/journal.pone.0117711
- Lister, P. D., Wolter, D. J., and Hanson, N. D. (2009). Antibacterial-resistant *Pseudomonas aeruginosa*: clinical impact and complex regulation of chromosomally encoded resistance mechanisms. *Clin. Microbiol. Rev.* 22, 582–610. doi: 10.1128/CMR.00040-09
- Luyt, C. E., Bréchet, N., Combes, A., Trouillet, J. L., and Chastre, J. (2013). Delivering antibiotics to the lungs of patients with ventilator-associated pneumonia: an update. *Expert Rev. Anti. Infect. Ther.* 11, 511–521. doi: 10.1586/eri.13.36
- Magill, S. S., Edwards, J. R., Bamberg, W., Beldavs, Z. G., Dumyati, G., Kainer, M. A., et al. (2014). Multistate point-prevalence survey of health care-associated infections. *N. Engl. J. Med.* 370, 1198–1208. doi: 10.1056/NEJMoa1306801
- Marchesi, J. R., and Ravel, J. (2015). The vocabulary of microbiome research: a proposal. *Microbiome* 3, 31. doi: 10.1186/s40168-015-0094-5

- McArthur, A. G., Wagelchner, N., Nizam, F., Yan, A., Azad, M. A., Baylay, A. J., et al. (2013). The comprehensive antibiotic resistance database. *Antimicrob. Agents Chemother.* 57, 3348–3357. doi: 10.1128/AAC.00419-13
- McLoughlin, K. S. (2011). Microarrays for pathogen detection and analysis. *Brief. Funct. Genomics* 10, 342–353. doi: 10.1093/bfgp/blr027
- Metzker, M. L. (2009). Sequencing technologies—the next generation. *Nat. Rev. Genet.* 11, 31–46. doi: 10.1038/nrg2626
- Morfopoulou, S., and Plagnol, V. (2015). Bayesian mixture analysis for metagenomic community profiling. *Bioinformatics* 31, 2930–2938. doi: 10.1093/bioinformatics/btv317
- Morris, A., Beck, J. M., Schloss, P. D., Campbell, T. B., Crothers, K., Curtis, J. L., et al. (2013). Comparison of the respiratory microbiome in healthy nonsmokers and smokers. *Am. J. Respir. Crit. Care Med.* 187, 1067–1075. doi: 10.1164/rccm.201210-1913OC
- Munoz, R., Yarza, P., Ludwig, W., Euzéby, J., Amann, R., Schleifer, K.-H., et al. (2011). Release LTPs104 of the all-species living tree. *Syst. Appl. Microbiol.* 34, 169–170. doi: 10.1016/j.syapm.2011.03.001
- Nhan, T.-X., Parienti, J.-J., Badiou, G., Leclercq, R., and Cattoir, V. (2012). Microbiological investigation and clinical significance of *Corynebacterium* spp. in respiratory specimens. *Diagn. Microbiol. Infect. Dis.* 74, 236–241. doi: 10.1016/j.diagmicrobio.2012.07.001
- Perez-Losada, M., Castro-Nallar, E., Bendall, M. L., Freishtat, R. J., and Crandall, K. A. (2015). Dual transcriptomic profiling of host and microbiota during health and disease in pediatric asthma. *PLoS ONE* 10:e0131819. doi: 10.1371/journal.pone.0131819
- Quail, M. A., Smith, M., Coupland, P., Otto, T. D., Harris, S. R., Connor, T. R., et al. (2012). A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. *BMC Genomics* 13:341. doi: 10.1186/1471-2164-13-341
- Rajendhran, J., and Gunasekaran, P. (2011). Microbial phylogeny and diversity: small subunit ribosomal RNA sequence analysis and beyond. *Microbiol. Res.* 166, 99–110. doi: 10.1016/j.micres.2010.02.003
- Rasko, D. A., Webster, D. R., Sahl, J. W., Bashir, A., Boisen, N., Scheut, F., et al. (2011). Origins of the *E. coli* strain causing an outbreak of hemolytic-uremic syndrome in Germany. *N. Engl. J. Med.* 365, 709–717. doi: 10.1056/NEJMoa1106920
- Rohde, H., Qin, J., Cui, Y., Li, D., Loman, N. J., Hentschke, M., et al. (2011). Open-source genomic analysis of Shiga-toxin-producing *E. coli* O104: H4. *N. Engl. J. Med.* 365, 718–724. doi: 10.1056/NEJMoa1107643
- Rose, H. D., Lenz, I. E., and Sheth, N. K. (1981). Meningococcal pneumonia: a source of nosocomial infection. *Arch. Intern. Med.* 141, 575–577. doi: 10.1001/archinte.1981.00340050027009
- Schmieder, R., and Edwards, R. (2011). Quality control and preprocessing of metagenomic datasets. *Bioinformatics* 27, 863–864. doi: 10.1093/bioinformatics/btr026
- Shetty, A., Barnes, R., Healy, B., and Groves, P. (2005). A case of sepsis caused by *Acidovorax*. *J. Infect.* 51, e171–e172. doi: 10.1016/j.jinf.2004.12.014
- Shinshkin, A. A., Giannoukos, G., Kucukural, A., Ciulla, D., Busby, M., Surka, C., et al. (2015). Simultaneous generation of many RNA-seq libraries in a single reaction. *Nat. Methods* 12, 323–325. doi: 10.1038/nmeth.3313
- Stamatakis, A. (2006). RAxML-VI-HP: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* 22, 2688–2690. doi: 10.1093/bioinformatics/btl446
- Toma, I., Siegel, M. O., Keiser, J., Yakovleva, A., Kim, A., Davenport, L., et al. (2014). Single-molecule long-read 16S sequencing to characterize the lung microbiome from mechanically ventilated patients with suspected pneumonia. *J. Clin. Microbiol.* 52, 3913–3921. doi: 10.1128/JCM.01678-14
- Turnbaugh, P. J., Ley, R. E., Hamady, M., Fraser-Liggett, C., Knight, R., and Gordon, J. I. (2007). The human microbiome project: exploring the microbial part of ourselves in a changing world. *Nature* 449, 804. doi: 10.1038/nature06244
- Ubukata, K., Nonoguchi, R., Matsushashi, M., and Konno, M. (1989). Expression and inducibility in *Staphylococcus aureus* of the *mecA* gene, which encodes a methicillin-resistant *S. aureus*-specific penicillin-binding protein. *J. Bacteriol.* 171, 2882–2885.
- Van Deuren, M., Brandtzaeg, P., and Van Der Meer, J. W. (2000). Update on meningococcal disease with emphasis on pathogenesis and clinical management. *Clin. Microbiol. Rev.* 13, 144–166. doi: 10.1128/CMR.13.1.144-166.2000
- Wang, B., Xu, J. S., Wang, C. X., Mi, Z. H., Pu, Y. P., Hui, M., et al. (2006). Antimicrobial susceptibility of *Neisseria gonorrhoeae* isolated in Jiangsu Province, China, with a focus on fluoroquinolone resistance. *J. Med. Microbiol.* 55, 1251–1255. doi: 10.1099/jmm.0.46401-0
- Weinstock, G. M. (2012). Genomic approaches to studying the human microbiota. *Nature* 489, 250–256. doi: 10.1038/nature11553
- Westermann, A. J., Gorski, S. A., and Vogel, J. (2012). Dual RNA-seq of pathogen and host. *Nat. Rev. Microbiol.* 10, 618–630. doi: 10.1038/nrmicro2852
- White, T. J., Bruns, T., Lee, S., and Taylor, J. W. (1990). “Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics,” in *PCR Protocols: A Guide to Methods and Applications*, eds M. A. Innis, D. H. Gelfand, J. J. Sninsky, and T. J. White (New York, NY: Academic Press, Inc.), 315–322.
- Willems, A., Falsen, E., Pot, B., Jantzen, E., Hoste, B., Vandamme, P., et al. (1990). *Acidovorax*, a new genus for *Pseudomonas facilis*, *Pseudomonas delafieldii*, E. Falsen (EF) group 13, EF group 16, and several clinical isolates, with the species *Acidovorax facilis* comb. nov., *Acidovorax delafieldii* comb. nov., and *Acidovorax temperans* sp. nov. *Int. J. Syst. Bacteriol.* 40, 384–398. doi: 10.1099/00207713-40-4-384
- Wilson, M. R., Naccache, S. N., Samayoa, E., Biagtan, M., Bashir, H., Yu, G., et al. (2014). Actionable diagnosis of neuroleptospirosis by next-generation sequencing. *N. Engl. J. Med.* 370, 2408–2417. doi: 10.1056/NEJMoa1401268
- Wright, G. D., and Poinar, H. (2012). Antibiotic resistance is ancient: implications for drug discovery. *Trends Microbiol.* 20, 157–159. doi: 10.1016/j.tim.2012.01.002
- Yang, S., and Rothman, R. E. (2004). PCR-based diagnostics for infectious diseases: uses, limitations, and future applications in acute-care settings. *Lancet Infect. Dis.* 4, 337–348. doi: 10.1016/S1473-3099(04)01044-8
- Yarza, P., Richter, M., Peplies, J., Euzéby, J., Amann, R., Schleifer, K.-H., et al. (2008). The all-species living tree project: a 16S rRNA-based phylogenetic tree of all sequenced type strains. *Syst. Appl. Microbiol.* 31, 241–250. doi: 10.1016/j.syapm.2008.07.001

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

EC, WJ, and KC have partial ownership in a pathogen diagnostics company, Aperiomics Inc., that uses metagenomic sequencing for pathogen detection. This company provided no payment or services associated with this study.

Copyright © 2016 Hilton, Castro-Nallar, Pérez-Losada, Toma, McCaffrey, Hoffman, Siegel, Simon, Johnson and Crandall. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.